illumına[®]

Designing and Ordering iSelect[®] HD Custom Genotyping Assays

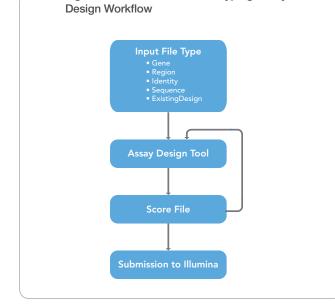
Guidelines for efficiently creating and ordering high quality custom Infinium[®] Genotyping Assays using the Illumina Assay Design Tool.

Introduction

The Illumina Infinium HD Assay is a powerful and widely used assay for highly multiplexed genotyping. Along with a broad portfolio of fixed content BeadArray[™] products, Illumina offers custom and semicustom genotyping panels deployed on various BeadChip formats. Researchers can design their own content panels for custom and semi-custom BeadChips with the assistance of the online Assay Design Tool (ADT) and Illumina scientists. The iSelect HD Custom Genotyping BeadChip can be completely custom designed to assay between 3,000 and 1,000,000 genotyping markers per sample in any species. The BeadChips come in the 24-sample format (3,000 to 90,000 attempted bead types), the 12-sample format (90,001 to 250,000 attempted bead types), or the 4-sample format (250,001 to 1,000,000 attempted bead types). iSelect+ combines custom content with the option of adding additional custom markers up to the limit of attempted bead types on the chosen sample format. The Human OmniExpress+ is a semi-custom genotyping BeadChip that combines standard human genome-wide content with the option of adding ~200K additional, custom markers.

Researchers create custom panels by selecting and submitting a list of requested loci to Illumina. To ensure successful assay development, this list is evaluated with the Assay Design Tool to select an initial assay panel including desired assays that are predicted to have a high likelihood of success. Metrics returned by ADT provide predicted success information, validation status, and minor allele frequencies from published studies.

Figure 1: iSelect Custom Genotyping Assay



different file types corresponding to the different methods for preliminary evaluation of custom SNP loci: Gene, Region, Identity, Sequence, and CNVRegion. This information is evaluated by ADT, and results are used to refine the panel definition. Each of the file type format options for ADT input and output are described with examples. Template files can be downloaded from iCom¹, or by contacting Illumina Technical Support². To get an iCom account, contact Illumina Technical Support.

This technical note describes the process of designing, analyzing,

and ordering a custom panel of markers. Researchers must submit

preliminary design files describing the custom panel using one of five

iSelect+ Infinium HD BeadChips

Illumina offers customers the ability to augment content on existing BeadChips with the latest advances from genome-wide association studies, next-generation whole-genome sequencing, and exome sequencing studies for variant confirmation, fine mapping, and target validation. This feature provides a significant benefit to customers who want to add newly discovered content to a custom array after the initial design period is completed.

The fully custom option (iSelect+), based on any content the researcher has the right to use (i.e., their own iSelect BeadChip, a commercial iSelect BeadChip, or an iSelect BeadChip shared amongst consortium users), allows further customization of content from any species. The number of additional markers that can be added is variable and depends on both the format of the BeadChip that content is being added to as well as the current complexity of the existing product. As reference, the 24-sample BeadChip can support up to 90,000 attempted bead types, the 12-sample BeadChip can support up to 250,000 attempted bead types, and the 4-sample BeadChip can support up to 1,000,000 attempted bead types.

Semi-Custom Infinium HD BeadChip

For human studies, the HumanOmniExpress+ BeadChip can be customized. The fixed content portion of the chip contains proven whole-genome SNP genotyping content, comprising ~733K markers. Researchers can expand upon this industry-leading genomic coverage by adding ~200,000 additional markers for specialized association studies with the semi-custom option.

Bead Type Definition

Depending on the type of SNP or marker being assayed, the Infinium HD Assay uses one of two probe (or bead type) designs: Infinium I or Infinium II. The Infinium II probe design, which stops at the base before the SNP of interest, uses only one probe per locus (i.e., one probe for both alleles). This probe design is suitable for the majority of loci in most organisms. Infinium I probe design is required for relatively less

common A/T and C/G SNPs and requires two probes (or bead types) per SNP because the probe stops at the base representing the SNP of interest (i.e., one probe for each of both alleles).

For ordering and pricing, custom BeadChip products are defined by their number of attempted bead types, not number of loci assayed. Only if a project is exclusively limited to loci using Infinium II designs (A/G, A/C, T/G, T/C SNPs; and all indels) are the number of markers equal to the number of bead types. The more markers ordered that require Infinium I design, the fewer total markers that can be attempted on a custom BeadChip.

For example, a researcher designing a custom iSelect BeadChip containing 9,500 Infinium II assays and 500 Infinium I assays would require a beadchip with a total of 10,500 attempted bead types: 9,500 Infinium II bead types (one probe per marker) plus 1,000 Infinium I bead types (two probes per marker).

Bead Type Success Rate

Illumina guarantees a bead type success rate of at least 80%. The bead type success rate is the proportion of the total number of attempted assays present on the final product. Manufacturing a custom BeadChip encompasses many steps, including synthesizing all required oligonucleotide probes, attaching probes to silica beads, combining the various bead types into the final bead pool, loading the bead pools onto the BeadChip, and decoding the bead identities on the BeadChip. While the manufacturing steps required to build custom BeadChips have been optimized, no step is 100% efficient. The manifest file supplied with BeadChips will contain all successful assays (i.e., those present on the final product), not all of the ordered assays. Illumina does not guarantee analytical success rates due to the custom nature of these projects.

Bead Pool and Product Shelf Life

Illumina guarantees the life of iSelect HD bead pools for 12 months from the date of manufacture. Illumina keeps custom bead pools in inventory for the entire 12 months to enable reorders of the same project. Once the 12 month shelf life has expired, any remaining bead pools are discarded. BeadChips are typically made to order and are not kept in inventory.

Illumina notifies customers of an impending bead pool expiration for product reorder of custom bead pools. In some cases, a bead pool may be kept past the 12 month expiration (by customer request), but the quality of the bead pool cannot be guaranteed, and Illumina reserves the right to not manufacture if the yield does not meet quality requirements.

Preliminary Input Files

To submit preliminary input files, researchers can use five different file types corresponding to the five different methods for preliminary evaluation of custom SNP loci: Gene, Region, Identity, Sequence, and CNVRegion. Additionally, probe designs from a previously ordered iSelect product or current Illumina standard product use the Existing-Design file type. After preliminary evaluation, ADT generates a Score file that can be used as an input file in subsequent rounds of evaluation or ordering (Figure 1). At this time, ADT returns only human sequences from Gene, Region, Identity, and CNVRegion input files. It is important to note that ADT only supports one build of the human genome at a time. Illumina keeps the supported version of the human genome current and gives users at least two weeks notice before switching to a new version. Illumina Technical Support Scientists can confirm which version of the human genome is currently supported.

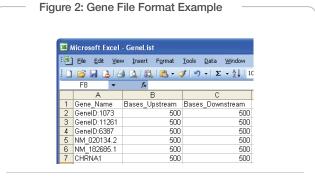
ADT supports full duplicate and repeat region checking for human assays. When designs are created for other species supported by ADT, varying levels of support for duplicate and repeat region checking will be given, depending on the availability of public sequence data. If ADT does not support the species, no duplicate and repeat region checking will be performed and customers should ensure their assays target unique sequences. For non-human species, Illumina Technical Support Scientists can provide information about the level of duplicate and repeat region support provided by ADT.

Input files may be created or edited with any text editor or spreadsheet program. However, before submitting them to ADT, files must be saved in a comma-separated values (*.csv) format. The examples provided in this document show files created in Microsoft Excel. Blank lines are not permitted in the data fields or between lines in the heading. The following formatting requirements must be followed precisely so ADT can properly evaluate your requests.

- Format is comma-separated values with a *.csv file extension. Since the input file format is comma-delimited, no commas may be used within the values.
- Each file type includes specific column headings for the data, as described below.
- File contains fewer than 1,000,000 markers or indels. If the number of markers exceeds this limit, the file must be split into smaller segments for serial ADT evaluation.
- If the file is submitted by email rather than on iCom, it must include a file header section. File header format is the same for all file types (Figure 8 and Table 7).

Gene

The Gene file type provides a method for querying all markers within a gene and in the regions upstream and downstream from the indicated gene. A Gene file allows for the interrogation of the currently supported build of the human genome using RefSeq NM accession ID (preferred) or HUGO identifiers. ADT maps these accession numbers



Example of properly formatted entries in a Gene file shown from Excel.

Table 1: Gene File Column Heading Descriptions

| Heading | Description | | | | |
|------------------|--|--|--|--|--|
| Gene_Name | Customer-supplied gene name. Can be a RefSeq accession ID or HUGO gene symbol. | | | | |
| Bases_Upstream | Number of bases to search upstream of the gene starting coordinate | | | | |
| Bases Downstream | Number of bases to search downstream of the gene starting coordinate | | | | |

to the human genome to identify gene regions. The sizes of upstream and downstream regions queried by ADT are specified by the user. Markers in overlapping gene regions will be listed in the Score output file only once, but will be annotated as being present in both regions in the Region_Description field. The column headings and description information shown in Table 1 must be provided in the Gene file. Figure 2 provides examples of proper Gene entries in Excel. A general guideline is that a Gene file with up to 600 genes and a range of 10 kb upstream and downstream will be within the marker limit.

Region

The Region file type enables researchers to select markers present between specified locations of a human chromosome. A Region file contains a list of regions in the human genome identified by chromosome and coordinate range that ADT will search and evaluate from among cataloged markers in a current Illumina-internal version of dbSNP. This internal database does not contain multinucleotide repeats (MNPs), microsatellites (simple sequence repeats), or markers with ambiguous or multiple localizations. Markers in overlapping regions will be listed in the Score output file only once, but will be annotated as being present in both regions in the Region_Description field. Since ADT limits output to 1,000,000 markers, submitting less than 60 Mb of regions per file is recommended. The column headings and description information shown in Table 2 must be provided in the Region input file. Figure 3 provides examples of properly formatted Region entries.

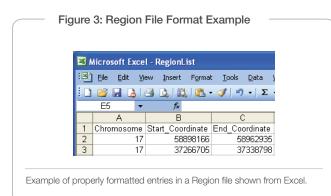


Table 2: Region File Column Heading Descriptions

Identity

Known markers described in the current version of dbSNP can be requested specifically using the Identity file type. A current internal version of dbSNP is the source for rs marker and flanking sequence data. The column headings and description information shown in Table 3 must be provided in the Identity input file. Figure 4 provides examples of properly formatted Identity entries.

Sequence

The Sequence file type allows researchers to evaluate markers from private databases or other sources, as well as from non-human species. The Locus_Name field is used to name sequences for easy identification. Locus_Name entries contained in this file must not begin with "rs," because that prefix designates rs ID names in the Illumina database and will trigger a database search.

To specify a SNP, put brackets around a polymorphic locus in the submitted sequence, and separate the two alleles with a forward slash (TGC[A/C]CCG). Similarly, to specify an indel, use a slash to separate a single minus sign indicating the deletion from the bases representing the insertion (TGC[-/AT]CCG). A minimum of 50 bp of sequence on either side of the SNP is required, but 60 bp is preferred. ADT will also accept IUPAC codes for degenerate bases in the flanking sequence and take these into consideration during design. If the Lowercase_ Weighting checkbox on the iCom submission form (or file header value) is unchecked, lowercase nucleotides will be considered for oligo design; if it is checked, then lowercase nucleotides are masked. In either case, an Illumina algorithm will identify repetitive or duplicated regions in the unmasked sequence. Since lowercasing in public databases is not a standard way to indicate masking, we recommend clearing the Lowercase_Weighting checkbox by default.

The column headings and description information shown in Table 4 must be provided in the Sequence input file. Figure 5 provides examples of properly formatted Sequence entries.

CNVRegion

The CNVRegion file type provides a method for researchers to select CNV markers present between specified locations of a human chromosome. CNV regions are typically analyzed by combining data from a set of polymorphic and nonpolymorphic probes. A CNVRegion file

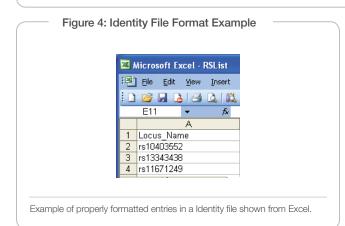
| Heading | Description | | | | |
|------------------|--|--|--|--|--|
| Chromosome | Chromosome on which the marker is located. Must be a valid chromosome for the species specified in the metadata. | | | | |
| Start_Coordinate | Non-negative integer indicating the start of the region of interest (inclusive) | | | | |
| End_Coordinate | Non-negative integer indicating the end of the region of interest (inclusive) | | | | |

| Table 3: Identity File Column Heading Description | |
|---|--|
|---|--|

| Heading | Description |
|------------|---|
| Locus_Name | rs number taken from dbSNP. Value must begin with "rs" (case insensitive) followed by an integer. |

Table 4: Sequence File Column Heading Descriptions

| Heading | Description | | | | | |
|---|---|--|--|--|--|--|
| Locus_Name | Customer-supplied unique locus name (cannot begin with rs or cg). Name must not contain any of the following characters: " $\% / = \$, @;', or space. Duplicate probe names are not allowed. | | | | | |
| Target_Type Must be "SNP" or "INDEL" (case insensitive) | | | | | | |
| Sequence | Limited to 10,000 bases. May only contain one bracketed SNP or indel. Output will be \leq 122 bases p | | | | | |
| Chromosome | Chromosome on which the marker is located. Must be a valid chromosome for the species being analyze Enter 0 if unknown. | | | | | |
| Coordinate | Chromosome coordinate of marker. Enter 0 if unknown. | | | | | |
| Genome_Build_Version | Version number supplied by customer. Otherwise, enter 0. | | | | | |
| Source | Identify the source of the sequence and annotation data. Must be completed. Enter unknown if no informa- tion is available | | | | | |
| Source_Version Source version number. Enter 0 if unknown. | | | | | | |
| Sequence_Orientation | Must be either "Forward" or "Reverse" (case insensitive) | | | | | |
| Plus Minus | Must be either "Plus" or "Minus" (case insensitive) | | | | | |



contains a list of regions in the human genome identified by a chromosome and coordinate range that are to be assayed for copy number variation (Table 5). The number of required probes in each region must also be specified.

CNVRegion files may optionally include a list of required polymorphic probes (Figure 6). When these markers are contained in a CNV region, they will be included in the set of markers for that region. The design module will intelligently select additional markers by balancing good probe scores with even spacing, resulting in a design consisting of high scoring and evenly spaced markers.

Using Previous Assay Designs

Illumina has created a separate method for conveniently ordering the same assays that were designed and used on a previous iSelect or current standard Illumina product. As described in Table 6 and shown in Figure 7, an ExistingDesign file only contains a list of the Illumina IDs (Ilmn_Ids) from the original design manifest.

Web Submitting via iCOM

Preliminary input files can be submitted to ADT for evaluation either directly via a web-based interface, or by emailing the file to an Illumina Technical Support Scientist who will submit the file to ADT. Submitting via the web is preferred since it provides rapid turnaround and 24-hour access. To submit a preliminary design file, log in to https:// iCom.illumina.com and select Prelim Assay Design Tool (ADT). The ADT interface allows users to enter necessary file type information and attach a *.csv formatted input file. After the file has been scored, an email notification is sent to the user.

Submitting via Email

Preliminary ADT input files submitted by email must include an additional file header section before the data entry section (Table 7). The format of the file header is common to all preliminary input file types. Email the entire *.csv file to techsupport@illumina.com. Figure 5: Sequence File Format Example

| 22 | Microsoft Excel - S | SequenceList | | | | | | | | |
|----|----------------------------------|--------------------|--|--------------------|------------|----------------------|--------|-----------------|----------------------|----------------|
| |] <u>F</u> ile <u>E</u> dit ⊻iew | Insert Format Iook | s <u>D</u> ata <u>W</u> indow <u>H</u> elp | Ado <u>b</u> e PDF | | | | | Type a question for | rhelp 👻 🗕 🗄 |
| |) 💕 🖬 💪 🔒 l | 🖪 🖪 🖓 🖏 🕹 | 🗈 🔁 • 🟈 🔊 • 🕅 | - 🧕 Σ - A | 🕴 👬 🛄 🦧 | 100% 🔹 🕜 💂 Arial | | • 10 • B | ⊻ ≣ ≣ ≣ ⊒ | - 🖄 - <u>A</u> |
| | K9 🔻 | fx | | | | | | | | |
| | A | B | C | D | E | F | G | Н | | J |
| 1 | Locus_Name | Target_Type | Sequence | Chromosome | Coordinate | Genome_Build_Version | Source | Source_Version | Sequence_Orientation | Plus_Minus |
| 2 | seq-rs10403552 | SNP | ATTCTCTTCCAGGGCT | 19 | 59900243 | 36 | dbSNP | 128 | Forward | Plus |
| 3 | seq-rs13343438 | SNP | cagtggcccaatctcagctca | 19 | 59751810 | 36 | dbSNP | 128 | Forward | Plus |
| 4 | seq-rs11671249 | SNP | tagaggtcattagcctaagtga | 19 | 59814896 | 36 | dbSNP | 128 | Forward | Plus |
| 5 | seq-rs34785303 | SNP | GAACCAGGCTGCTAA | 19 | 59976309 | 36 | dbSNP | 128 | Forward | Plus |
| 8 | sea-rs35835276 | SNP | CCCAGCCTGGGTTTTC | 19 | 59937831 | 36 | dbSNP | 128 | Forward | Plus |

Example of properly formatted entries in a Sequence file shown from Excel.

Score Output File

If preliminary input files are submitted to ADT via iCom, an email notification is sent when scoring is complete, often within an hour. The results are returned as a Score file that can be downloaded from iCom on the Prelim Assay Design files page. If an input file is submitted via email to Technical Support, an Illumina scientist will submit the file to ADT for processing. ADT generates the Score output file, which is returned to the customer by email or secure FTP in 1–2 days. The Score file provides information for filtering and selecting their marker lists, as described in the following section.

The Score file header section includes additional summary information, such as the total number of markers in the file. This is further broken down into numbers of markers in each of three Normalization_Bins: A, B, and C. Bin C assays are Infinium II designs. Bin A and B assays are Infinium I designs. Infinium I assays are classified into one of these two bins based on the sequence context (the base pair after the locus) and strand that the assay is designed against. For a final order, if there are any loci in a given normalization bin, there must be at least 100 loci in that bin to ensure normalization functions as expected. If Bin A or Bin B needs to be supplemented to reach 100, submit additional A/T or C/G SNPs to ADT for scoring. ADT will report whether each of these SNPs are in Bin A or Bin B. The appropriate assays can then be added to the original design as needed.

The Number of Bead Types value is important for ordering, since iSelect BeadChip orders are priced based on the number of attempted

Figure 6: CNVRegion File Format Example

| | Microsoft Excel - WG | GT_ | CNVRegio | on-1 [R | ead-0 | nly] | | | l | |
|----|----------------------|------|-----------|---------------|--------------|-----------------|-------|--------------|-------|-------|
| 1 | Eile Edit View Ins | sert | Format | <u>T</u> ools | <u>D</u> ata | <u>W</u> indow | Help | Ado <u>t</u> | e PDF | - |
| ED | 💕 🔒 👌 🎒 🛕 | 1 62 | L 🖻 - 🤇 | 19 | - Σ | - <u>2</u> ↓ [] | 10 |)0% · | • 0 | 21 |
| | F10 - | f×. | | | | | | _ | | |
| | A | | В | | | С | | | D | |
| 1 | Chromosome | ę | Start Coo | rdinate | End (| Coordinat | te Pr | obes F | Per R | egior |
| 2 | | 17 | - 58 | 898166 | _ | 5896293 | 35 | _ | | |
| 3 | | 17 | 37 | 266705 | | 3733879 | 98 | | | |
| 4 | | 1 | 54 | 776489 | | 548830 | 35 | | | |
| 5 | | 3 | 188 | 033157 | | 1880689 | 46 | | | |
| 6 | | 1 | 201 | 166576 | | 2012040 | 34 | | | |
| 7 | [MISC] | | | | | | | | | |
| 8 | Required_Locus_Nan | ne | | | | | | | | |
| 9 | rs10403552 | | | | | | | | | |
| 10 | rs13343438 | | | | | | | | | |
| 11 | rs11671249 | | | | | | | | | |

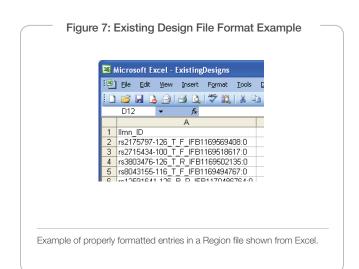
bead types. The number of bead types may be different than the number of assays because Infinium I assays require two bead types per markers and Infinium II assays only require one bead type per marker.

Following the Score file header section, detailed information for each marker is listed in the data section. All Score column headers are described in Table 8. Important performance values are also presented for each SNP. The Final_Score indicates the expected success for designing a given assay, and may be supplemented with Failure_Codes for further information. Validation status is also indicated to provide even greater confidence in design success. To assist researchers in ordering the most applicable markers for their studies, minor allele frequencies (MAFs) in several populations are provided for SNPs when available from dbSNP. MAF from the largest study is reported, and is qualified based on peer-reviewed publication, study design and size, and verified results.

Filtering and Selecting Custom List

In addition to being an output file format, Score files can be used as input files to ADT. Thus, users can create a filtered or edited output file (with markers removed or added) for iterative ADT analysis while determining optimal loci. Markers identified using more than one input search method (e.g., Gene, Region, Identity, Sequence, CNVRegion, or ExistingDesign) can be combined as one Score file and resubmitted to ADT as an input file for evaluation as a single product.

Illumina recommends applying the following criteria for discriminating



marker lists to create a successful product that achieves the scientific aims of the experiment and has the highest chances of generating meaningful results.

- 1. Remove markers that cannot be ordered (error codes in the 101–199 range).
- 2. Consider research requirements (e.g., tags, spacing, or MAF).
- 3. When appropriate, favor Infinium-validated markers, since they have the highest chance of converting into functional assays.
- 4. Use two-hit validated markers based on Final_Score. Higher Final_Scores are preferred.
- 5. Do not be hindered by proximity limits for Infinium assays. Markers can be close to each other without potential interference.
- Submit a number of bead types that is within the required range. Human OmniExpress+ require 1,000 to 200,000 attempted bead types. iSelect HD BeadChips can accommodate 3,000 to 1,000,000 attempted bead types.

Final Score File

After selecting SNPs that meet your research criteria and the ADT criteria, a final Score file must be created to place an order. A preliminary Score file is converted to a final Score file by the completion of four existing header rows: Design_Iteration, Scale (Number_of_Tubes), Purchase_Order_Number, and Product_Name. It is important to ensure that the Number of Bead Types (not SNPs) value in the final file matches the number on the corresponding quotation or contract.

Figure 8: Examples of the File Header Section Required Only For Email Submission

| 9 | ∣ <u>E</u> ile <u>E</u> dit ⊻iew Insert F <u>o</u> r | mat <u>T</u> ools | <u>D</u> ata | <u>W</u> indow <u>H</u> | <u>i</u> elp Ado <u>b</u> e PC |)F |
|----|--|-------------------|--------------|-------------------------|--------------------------------|---------|
| 1 | 🗃 🖬 👌 🖨 🛕 🛍 🕻 | L - 🛷 🖂 | - Σ | - 2↓ 100 | 1% 🝷 🕜 | ** Ŧ |
| | F18 🔻 🏂 | | | | | |
| | A | В | C | D | E | |
| 1 | [heading] | | | | | |
| 2 | Customer_Name | Jane Doe | | | | |
| 3 | Company_Name | BigBio Inc. | | | | |
| 4 | Company_Address1 | 1234 Biote | ch Way | (| | |
| 5 | Company_Address2 | | | | | |
| 6 | City | San Diego | | | | |
| 7 | State/Province | CA | | | | |
| 8 | Postal_Code | 92121 | | | | |
| 9 | Country | USA | | | | |
| | Phone_Number | 858-202-1234 | | | | |
| 11 | Email_Address | jdoe@illum | | n | | |
| 12 | | SNPsforPr | | | | |
| 13 | Order_Comments | | nent isn | t used by a | iny ILMN soft | war |
| 14 | = 21 | WGGT | | | | _ |
| 15 | | 0 | | | | _ |
| 16 | Format_Type | Score | | | | |

Examples of properly formatted entries in the header of a Score file shown from Excel. This header is only required for email submissions, and is formatted the same for any preliminary file format.

Ordering

To place an order, submit the final Score file on the iCom website (https://iCom.illumina.com, Figure 9). Options for reordering an existing iSelect panel or obtaining customized project pricing information before placing an order are also available on iCom.

Upload iSelect HD Design Files

To upload files to iCom, follow these steps:

- 1. On the iCom welcome page, select **Orders | Upload iSelect HD Files**.
- 2. Click Upload Final Score File.
- 3. Select the File Type (Score or ExistingDesign).
- 4. Select the Assay Type (WGGT).
- 5. Select the **Species**.
- Enter the Number of Assays, which is the number of different markers or variants (e.g., SNPs or CNVs) being interrogated in the custom project, not the number of bead types.
- 7. Enter the **Description** (user-defined and optional).
- 8. Enter the location and file name of the target file or click **Browse** to navigate to the file.
- 9. Click Upload and Continue with Validation. The uploaded file will then be listed in the iSelect Files table, with an initial status of "In Scoring." Within two days, ADT scoring will be completed, and an email notification will be sent. At that point, the file status will be updated to "Ready to Order" or "Warning." See Table 9 for a description of possible warning codes.
- 10. After a file is uploaded and successfully validated, it can be used for product order placement by selecting the checkbox next to the file and clicking **Continue to Project Pricing**.

iSelect HD Project Pricing Calculator

The Project Pricing Calculator can be used at any time. It opens automatically when you select **Continue to Project Pricing** for a design file as described above. To get project pricing information at any other time, go to the iCom web page and select **Tools** | **iSelect Project Pricing Calculator** or **Order Now**. On the next screen, click **Infinium Genotyping** | **Project Wizard**.

- 1. Select the **Product Type:** iSelect, iSelect+, or HumanOmniExpress+ (Figure 10).
- 2. If ordering iSelect+, select the BeadChip you wish to supplement in the **Existing Content** field.
- 3. Enter the number of samples between 1,152 and 116,352 in multiples of 48.
- 4. Do one of the following:
 - To determine pricing based on a design file, select the appropriate file in the Final iSelect File section.

Table 5: CNVRegion File Column Heading Description

| Heading | Description | | | | |
|----------------------|--|--|--|--|--|
| Chromosome | Chromosome on which the marker is located. Must be a valid chromosome for the species specified in the metadata. | | | | |
| Start_Coordinate | Non-negative integer indicating the start of the region of interest (inclusive) | | | | |
| End_Coordinate | Non-negative integer indicating the end of the region of interest (inclusive) | | | | |
| Probes_Per_Region | The number of probe designs that should be returned for the region of interest | | | | |
| Required_Locus_Name* | RS number taken from dbSNP; value must begin with "rs" (case insensitive) followed by an integer | | | | |

* Optionally included to specify required polymorphic probes

| Table 6: Colum | n Headings for Existing Design File |
|----------------|---|
| Heading | Description |
| ILMN_ID | ID value of an ExistingDesign file, which must map to a design with the same species and assay type as specified in the metadata. |

Table 7: File Headings for ADT Input File Required Only for Email Submission

| Heading | Description | Required | | |
|---------------------|--|----------|--|--|
| Customer_Name | Name of person submitting the ADT file | Yes | | |
| Company_Name | Company name (no commas) | Yes | | |
| Company_Address1 | mpany_Address1 Line 1 of customer's address | | | |
| Company_Address2 | Line 2 of customer's address | No | | |
| City | Customer's city | No | | |
| State/Province | Customer's state or province | No | | |
| Postal_Code | Customer's postal code | No | | |
| Country | Customer's country | No | | |
| Phone_Number | Customer's phone number | No | | |
| Fax_Number | Customer's fax number | No | | |
| Email_Address | Customer's email address | Yes | | |
| Order_Description | Description of work (30 character limit) | No | | |
| Order_Comments | Additional comments | No | | |
| Platform_Type | WGGT | Yes | | |
| Number_of_Assays | Number of loci in file (enter 0 for GeneList and RegionList) | No | | |
| Lowercase_Weighting | 0 for no masking or 1 to mask lower cased characters | No | | |
| Format_Type | Gene, Region, Sequence, Identity, ExistingDesign, or Score | Yes | | |
| Species | Species name | Yes | | |

Table 8: Score File Column Headers

| Heading | Description | | | | |
|---------------------------------|--|--|--|--|--|
| Locus_Name | rs number or customer's unique name | | | | |
| Sequence | The bracketed site identified by the Locus_Name with > 50 bases of flanking sequence | | | | |
| Genome_Build_Version | Genome build that will be queried. Contact Technical Support ² for the currently supported build | | | | |
| Chromosome | Chromosome on which the marker is located. Must be a valid chromosome for the species being analyzed. Enter 0 if unknown. | | | | |
| Coordinate | Chromosome coordinate of marker. Enter 0 if unknown | | | | |
| Source | Identify the source of the sequence and annotation data. Enter unknown if no information is available | | | | |
| Source_Version | Source version number or 0 if unknown | | | | |
| Sequence_Orientation | Must contain one of the following three values: forward, reverse, or unknown. Information is customer- supplied and is not validated. | | | | |
| Region_Description | Description of the region of interest | | | | |
| Final_Score | Ranges from 0-1.1, with higher values reflecting greater ability to design a successful assay | | | | |
| Failure_Codes | If applicable, reasons why a successful assay at this marker locus is unlikely (Table 9) | | | | |
| Validation_Class | Numerical representation of validation_bin (Table 10) | | | | |
| Validation_Bin | Manner in which designed assays have been validated (Table 10) | | | | |
| MAF_Caucasian | | | | | |
| Chr_Count_Caucasian | | | | | |
| Study_Name_Caucasian | | | | | |
| MAF_African | | | | | |
| Chr_Count_African | | | | | |
| Study_Name_African | | | | | |
| MAF_African_American | Minor allele frequency from the largest peer-reviewed study conducted, the study size in terms of num- | | | | |
| Chr_Count_African_ American | ber of chromosomes, and the study type. Data are retrieved from dbSNP for each population: Caucasian | | | | |
| Study_Name_African_ American | African African-American | | | | |
| MAF_Japanese | Han Chinese | | | | |
| ChrCount_Japanese | Japanese | | | | |
| Study_Japanese | Unknown | | | | |
| MAF_Chinese | | | | | |
| Chr_Count_Chinese | | | | | |
| Study_Name_Chinese | | | | | |
| MAF_Other | | | | | |
| Chr_Count_Other | | | | | |
| Study_Name_Other | | | | | |
| Normalization_Bin | A, B, or C | | | | |
| Bead_Types/Assay | 1 for Infinium II or 2 for Infinium I | | | | |
| Assay_Type | Infinium I or Infinium II | | | | |
| ILMN_ID | Unique identifier assigned by ADT for the designed assay | | | | |
| Gene_ID* | Gene ID number from NCBI | | | | |
| Gene_symbol | HUGO identifier | | | | |
| Accession* | RefSeq Accession number | | | | |
| Location* | Structural location of the marker: intron, coding, flanking_5UTR, flanking_3UTR, 5UTR, 3UTR, UTR | | | | |

Table 8, (Continued) Score File Column Headers

| Heading | Description |
|----------------------------|--|
| Location_relative_to_gene* | If the marker does not fall within an exon, the value is the actual base pair distance from gene start. The absolute value of this number is the distance to the closest transcript. The negative sign is a formatting symbol and is not meant to imply strand or direction. If the SNP is within an exon, two values separated by a '/' are gvien, which represent distances to the exon-intron boundaries. |
| Coding_status* | NONSYN or SYNON. If the marker falls within an exon, this field notes a synonymous or nonsynony- mous amino acid change. |
| Amino_acid_change* | If the marker falls within an exon, this field notes the actual change to the amino acid, followed by the GenBank protein sequence used in numbering the change |
| ld_with_mouse* | Ratio of identical bases within 60 bp of flanking sequence compared to mouse sequence that has been aligned with the homologous human sequence and covers the marker in question |
| Phast_conservation* | Metric used by the UCSC Genome Browser to identify highly conserved markers among species |
| Design Date | Date of design |

*Additional gene annotation only in Score output file from submitted GeneList, RSList, and RegionList files

- To determine the pricing based on an assumed number of bead types rather than a design file, enter the number of bead types for pricing calculation.
- 5. Click Calculate Price.
- 6. To place an order based on the selected design file, enter the product name and click **Preview this iSelect Product**.
- 7. Click Add to Cart.

Reordering an Existing Bead Pool or Product

To place an order based on an existing product or custom bead pool, click **Order Now**, and then click **Reorder iSelect**. Reorders have a minimum order requirement of 288 samples, and must be ordered in multiples of 48. Only non-expired bead pools are available for reordering. Bead pools expire after 12 months. Unused final design files are removed 90 days after being uploaded. If a design file has been used to place an BeadChip order, that design file is never removed.

To reorder a custom product identical to one that has already been ordered:

- 1. Select the desired custom product in the list.
- 2. Click Reorder to go to the iSelect Project Calculator.
- In the iSelect HD Project Calculator, select the sample number desired.
- 4. Click **Preview** to view the order summary.
- 5. Click **Add to Cart** to add the reordered product to the cart, and complete the order as described in the next section.

Completing an Order

Once all desired orders are added to the shopping cart, click **Continue to Checkout**. The shopping cart is automatically displayed after clicking **Add to Cart**; otherwise click **View Cart**. During the checkout process, shipping and payment information is required. An initial shipping date and partial shipment request can be entered before final submission. After clicking **Submit**, the order will be sent to Illumina Customer Solutions for review. If needed, customers can make modifications to an order by emailing orders@illumina.com during the 48 hour review period.

Shipping Schedule

Shipping schedules can be defined after an order is placed. To specify a shipping schedule, fill out the Ship Schedule Form (available from a link in the order confirmation or directly from https://icom.illumina.com/ EAI/ShipScheduleForm.xls) and email it within two weeks of the order confirmation date to orders@illumina.com. Otherwise, orders may ship complete or according to a default schedule. The first shipment of custom BeadChips should arrive within 8–10 weeks after order confirmation.

Projects with fewer than 12,000 samples must be shipped complete within 6 months. Projects of 1,152 to 2,304 samples have a maximum of two shipments, 2,305 to 5,760 samples have a maximum of three shipments, and 5,761 to 11,999 have a maximum of four shipments. Projects of over 12,000 samples must be shipped within 12 months, and a weekly shipping schedule can be arranged.

Ordering by Email

To place an iSelect HD BeadChip order by email, submit the final Score file and an official purchase order to orders@illumina.com. Illumina will send a confirmation and contact you if essential information is missing or if it contains unorderable designs.

Summary

Custom Infinium DNA Analysis products by Illumina allow researchers to create assays tailored to their specific needs for targeted region genotyping or fine-mapping of candidate disease association regions. The ADT is a simple and powerful tool for evaluating individual loci and creating the successful custom BeadChips for genotyping. By following the guidelines in this technical note, researchers can ensure that their orders are designed and placed quickly and easily.

References

- 1. https://icom.illumina.com/
- To contact Technical Support, send email to techsupport@illumina.com or call 1.800.809.4566.

Table 9: List of ADT Failure Codes

| 01 | Flanking sequence is too short |
|----------|---|
| 102 | Formatting error. Variant must match the format [A/B]. Possible causes: |
| | Spaces in submitted sequence |
| | More than one set of brackets in sequence |
| | Missing brackets around variant |
| | Variant alleles not separated by "/" |
| 103 | TOP/BOT strand cannot be determined due to low sequence complexity |
| 104 | Variant is not appropriate for Illumina platform. Possible causes: |
| | Tri- or quad-allelic variant |
| | Variant contains characters other than A, G, C, and T |
| 106 | Degenerate nucleotide(s) in assay design region (e.g., W, R, S, N) |
| 109 | Indels are not supported in the Infinium I Assay |
| Warnings | (designable) |
| 301 | Variant in duplicated/repetitive region |
| 399 | Multiple contributing issues |

Table 10: Validation Status Descriptions

| /alidation_Bin | Validation_Class | Description | | | | | |
|-----------------------------------|------------------|---|--|--|--|--|--|
| NonValidated | 1 | Locus has been seen in only one method or population. Even if it has a high design score, there is an increased chance that it is monomorphic. | | | | | |
| OneKGenomeValidated | 100 | Locus has been sequenced in the 1000 Genomes project | | | | | |
| TwoHitValidated | 110 | Both alleles have been seen in two independent methods and populations | | | | | |
| HapMapValidated | 120 | Locus has been genotyped by the HapMap project | | | | | |
| TwoHit_OneKGenomeValidated | 200 | Both alleles have been seen in two independent methods and populations. Locus has been sequenced in the 1000 Genomes project. | | | | | |
| HapMap_OneKGenomeValidated | 210 | Locus has been genotyped by the HapMap project. Locus has been se- quenced in the 1000 Genomes project. | | | | | |
| TwoHit_HapMapValidated | 220 | Locus has been genotyped by the HapMap project. Both alleles have been seen in two independent methods and populations. | | | | | |
| TwoHit_HapMap_OneKGenomeValidated | 300 | Locus has been genotyped by the HapMap project. Both alleles have been seen in two independent methods and populations. Locus has been se- quenced in the 1000 Genomes project. | | | | | |
| InfiniumValidated | 910 | Design has been previously designed and successfully generated polymorphi results using the Infinium assay | | | | | |

Figure 9: Upload iSelect Design Files

| Home / Custom | Orders / iSelect Files / Upload ISelect 800.809.4566 |
|----------------|---|
| ISFLECT LINKS | Upload iSelect File |
| Custom Home | Please note that all Final files must go through preliminary scoring in order to be uploaded here in the proper format. Click here for scoring templates and a tutorial on creating outsom (Selects, For more information on probe design, please refer to our |
| ISelect Chips | iSelect Design Tech Note available for download from iCom. |
| | Please, upload the data file containing your iSelect data. Use the "Browse" button to identify the location of your data file. If the file name is longer than 50 characters, it will be truncated. |
| Uploaded Files | Fule Type* Score |
| Upload Prelim | Assay Type* Infinium Genetyping |
| Upload Final | Species* Human v Number of Loci* 138000 |
| Calculator | Description* Final Design File |
| ADT Templates | |
| Status Codes | Select Die |
| | Browse |
| My Account | Upload for Validation |
| My Orders | |
| My Saved Carts | |
| My Lists | |
| iCom Home | |

| illumina' | | | | | reenal Sequen | Help Lopout |
|--------------------|--|--------------------|------------------|---------------|---------------|-------------|
| APPLICATIONS SH | TEMS SERVICES SCIENCE | SUPPORT | COMPANY | | | |
| Home / Custom Orde | rs / iSelect Files / iSelect Calculat | or | | View Cart | | 00.809.4566 |
| Products | Project Calculator | | | | | |
| ISTLUCT LINKS | Select Product Type* C idelect | | | | 111 | |
| Custom Home | Galacter + shall in Uhap C Human Onno-Express+ shall in Uhap Existing Contexts (2011-1-4 * | | | | 1000 | |
| Uploaded Files | | | | | <u></u> | |
| Upload Prelim | | | | Lation | | |
| Uplead Final | | ober of samples ro | unded up to mini | mum order qua | shity of 1152 | |
| Calculator > | Select pricing method* @ calent Fles when is thin? C Enter Number of Beadways | | | | | |
| ADT Templates | | | | | | |
| Status Codes | Final Belect Files Select File Name | Species | Assay Count | Bead Types | mounded | Designation |
| Ply itelect Chips | national university_wggt_0927649. Beaut Point Reservicer | Home Sapiena | 5222 | 6000 | 12/17/2010 | |
| | C state Like_canine_88292_ | Other | 73370 | 74000 | 12/20/2010 | N/A |
| QUICK LINKS | Bead Pool Reorder | | | | | |
| Home | | | | | | |
| By Account | | | | | = | Set Pricing |
| Downloads | | | | | | |
| Custern Orders | | | | | | |
| Quick Order | | | | | | |
| Hy Orders | | | | | | |

Additional Information

Visit our website or contact us to learn more about iSelect Custom Infinium products from Illumina.

Illumina, Inc. • 9885 Towne Centre Drive, San Diego, CA 92121 USA • 1.800.809.4566 toll-free • 1.858.202.4566 tel • techsupport@illumina.com • illumina.com

FOR RESEARCH USE ONLY

© 2011 Illumina, Inc. All rights reserved.

Illumina, illuminaDx, Solexa, Making Sense Out of Life, Oligator, Sentrix, GoldenGate, GoldenGate Indexing, DASL, BeadArray, Array of Arrays, Infinium, BeadXpress, VeraCode, IntelliHyb, iSelect, CSPro, GenomeStudio, Genetic Energy, HiSeq, HiScan, TruSeq, Eco, and MiSeq are registered trademarks or trademarks of Illumina, Inc. All other brands and names contained herein are the property of their respective owners. Pub. No. 370-2007-021 Current as of 25 January 2011

